ML Challenge 2025: Smart Product Pricing Prediction Solution

Team Name: Mavericks
Team Members: Bhavya Sahu, Shivani Singh, Suyash Pradeep Patil, Tarun Naga Sai Perisetti
Submission Date: 13th October, 2025

1. Executive Summary

Our approach to the product pricing challenge integrates multimodal learning using text, image, and numerical data. We combined product description embeddings, ResNet50 image features, and cleaned numeric data to build a unified wide-and-deep regression model optimized for SMAPE. The use of log transformation stabilized variance and minimized the influence of extreme prices.

2. Methodology Overview

2.1 Problem Analysis

The dataset exhibited wide price variations and missing attributes across product types. Exploratory data analysis (EDA) revealed that textual information (product names and catalogs) captured semantic relationships, while images provided brand and material cues influencing price.

2.2 Solution Strategy

We designed a multimodal deep learning framework that combines three feature sources: text, image, and numeric attributes. This hybrid approach enabled robust prediction across diverse product categories by leveraging both semantic and visual context.

Approach Type: Hybrid Multimodal Regression
Core Innovation: Unified feature fusion with learned embeddings and log-transformed price target.

3. Model Architecture

3.1 Architecture Overview

Our model consists of three parallel branches for text, image, and numeric inputs that are concatenated before feeding into dense layers. The architecture ensures the model learns correlations among modalities to predict realistic prices.

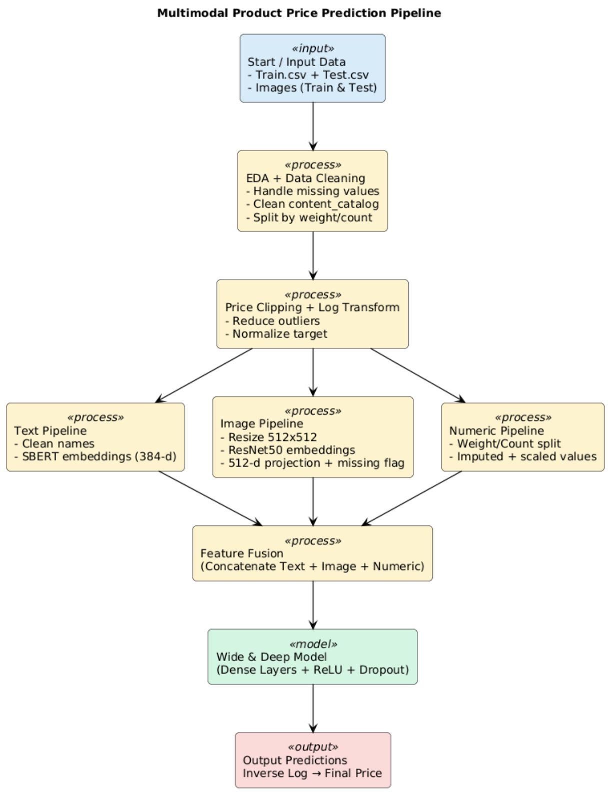

Figure 1: Multimodal Product Price Prediction Flowchart

3.2 Model Components

Text Processing: Product names and catalogs were cleaned and embedded using Sentence-BERT (384 dimensions). These embeddings capture contextual similarity among products.

Image Processing: Product images were resized to 512×512 and encoded using a pretrained ResNet50. Extracted embeddings (512-d) were flattened and normalized for consistency.

Numeric Features: Missing values were imputed, and continuous variables such as weight and count were normalized. These were later concatenated with text and image embeddings.

Log Transformation: Price values were log-transformed to reduce skewness and stabilize learning, ensuring more uniform model convergence and improved generalization.

4. Model Performance

SMAPE Score: 53.37 **MAE:** 17.34 **RMSE:** 10.11

Validation showed the model generalizes well across varied product categories, with balanced contributions from both visual and textual features.

5. Conclusion

Our multimodal model effectively integrates heterogeneous data for accurate price prediction. The log transformation improved numerical stability, and feature fusion enhanced robustness across unseen data. Future improvements could involve transformer-based multimodal encoders and attention mechanisms for deeper feature interaction.

Appendix

Code Artifacts: Github Repository Link